Towards image-guided pancreas and biliary endoscopy: automatic multi-organ segmentation on abdominal CT with dense dilated networks

Eli Gibson[1], Francesco Giganti[2,3], Yipeng Hu[1], Ester Bonmati[1], Steve Bandula[4], Kurinchi Gurusamy[3], Brian R. Davidson[3], Stephen P. Pereira[5], Matthew J. Clarkson[1], and Dean C. Barratt[1]

[1] UCL Centre for Medical Image Computing, Department of Medical Physics & Biomedical Engineering, University College London, UK
[2] Department of Radiology, University College Hospital Trust, UK
[3] Division of Surgery and Interventional Science, University College London, UK
[4] UCL Centre for Medical Imaging, University College London, UK
[5] Institute for Liver and Digestive Health, University College London, UK

Abstract. Segmentation of anatomy on abdominal CT enables patient-specific image guidance in clinical endoscopic procedures and in endoscopy training. Because robust interpatient registration of abdominal images is necessary for existing multi-atlas- and statistical-shape-model-based segmentations, but remains challenging, there is a need for automated multi-organ segmentation that does not rely on registration. We present a deep-learning-based algorithm for segmenting the liver, pancreas, stomach, and esophagus using dilated convolution units with dense skip connections and a new spatial prior. The algorithm was evaluated with an 8-fold cross-validation and compared to a joint-label-fusion-based segmentation based on Dice scores and boundary distances. The proposed algorithm yielded more accurate segmentations than the joint-label-fusion-based algorithm for the pancreas (median Dice scores 66 vs 37), stomach (83 vs 72) and esophagus (73 vs 54) and marginally less accurate segmentation for the liver (92 vs 93). We conclude that dilated convolutional networks with dense skip connections can segment the liver, pancreas, stomach and esophagus from abdominal CT without image registration and have the potential to support image-guided navigation in gastrointestinal endoscopy procedures.

Keywords: Abdominal CT · Segmentation · Pancreas · Stomach · Liver · Esophagus · Deep learning

1 Introduction

In interventional endoscopy for pancreatic and biliary diseases, navigating the endoscope to specific gastrointestinal (GI) positions and orientations is critical for both diagnosis and treatment [4]. The endoscope's small field of view and lack of visual orientation cues make this navigation task challenging, particularly for novice endoscopists [4]. Image-guidance can support intraprocedural navigation tasks and endoscopy training by revealing the wider anatomical context.

As the shapes and positions of abdominal organs can vary widely between patients, *patient-specific* anatomical models of the GI tract and surrounding organs should enable more accurate alignment with intra-procedural imaging and may improve performance. These models can be generated from segmented abdominal CT; however, integration into clinical workflows is only practical if the segmentation can be highly automated.

Multi-organ segmentation has been the subject of extensive study. The most common approaches, statistical shape models [2,9] and multi-atlas label fusion [16,9,14,13,12], rely on registration to establish anatomical correspondence. However, interpatient image registration is less accurate for abdominal imaging than for other anatomical sites (e.g. brain), due to highly variable anatomy [14]. Thus, there is a need for automated multi-organ segmentation that does not rely on registration or shape-model fitting.

Deep-learning-based fully convolutional networks offer an approach to segment anatomy from voxel-based features directly. These networks have been successfully applied to segment individual organs from medical images, such as prostate [8] and pancreas [11]. They have also shown promise in abdominal CT for multi-organ segmentation [5] of the liver, spleen and kidney.

In this study, we present a fully convolutional network to segment the liver, pancreas, stomach and esophagus from abdominal CT. Such segmentations enable patient-specific 3D modelling of the GI tract and surrounding anatomy, providing a navigational reference for endoscopists. The network is trained and evaluated on 72 abdominal CT images from two centres, and directly compared to an existing approach based on multi-atlas image registration and label fusion.

2   Methods

Imaging Abdominal CT images from two datasets were used in this study: 42 images from the Cancer Imaging Archive Pancreas-CT data set [3,11,10] and 30 images from the 'Beyond the Cranial Vault' segmentation challenge (doi:10.7303/syn3193805). For the latter, manual reference segmentations of liver, pancreas, stomach and esophagus were available. For the former, manual reference segmentations of the pancreas were available and the liver, stomach and esophagus were interactively segmented using Matlab 2015b and ITK-SNAP 3.2 (`http://itksnap.com`), under the supervision of a board-certified radiologist with 8 years of experience in gastrointestinal CT and MRI image interpretation. Images were cropped to the ribcage transversely and to the extent of the segmented organs in the inferior-superior direction.

Dense dilated convolutional network segmentation The proposed segmentation used a fully-convolutional neural network with dilated convolution units with dense skip connections, described below and illustrated in Figure 1a.

Dilated convolutions [17] use sparse convolution kernels allowing a large kernel spatial extent without increasing the number of learned parameters. Compared to using a cascade of downsampling layers [8], dilated convolutions main-

tain a high-resolution representation of non-local non-linear image features deeper in the network. This is particularly important in organ segmentation where thin structures (e.g. thin liver-adjacent stomach walls) must be inferred based on high-level information (e.g. adjacent stomach and liver tissue). Dilated convolutions can be implemented efficiently by reordering input data in memory (Fig. 1b) and convolving with a corresponding non-sparse convolution, leveraging efficient algorithms and hardware-support. Each convolutional layer is followed by a batch normalization and a rectifier linear unit defining a structure referred to hereafter as a convolutional unit.

The network uses an initial convolutional feature layer, 8 convolutional units, a segmentation layer and a spatial prior. The feature layer outputs 25 outputs with 5^3 convolutional kernels and stride 2, and a downsampled image. The convolutional units used 3^3 kernels with dilation scales of 1, 1, 2, 2, 4, 4, 2, and 1, each outputting 20 feature maps. The segmentation layer outputs one logit map for each label (liver, pancreas, stomach, esophagus, and other).

The convolutional units have dense skip connections [6]; i.e., the input to each unit is the concatenated output of all previous units. This enables efficient use of intermediate features as intermediate layers do not need to re-encode information from previous layers. Additionally, like shortcut layers in residual networks, the dense skip connections allow effective propagation of gradients through the network and combine multiple networks depths in the same network.

Finally, we introduce a new spatial prior map, added to the segmentation unit output. Spatial priors are more suited to medical images than natural images because medical images are commonly acquired in standard anatomically aligned views. The map comprised a 12^3 block of trainable parameters which was upsampled and added to the logit maps. This is analogous to the logit maps representing the log-probability of the input given the class label, the spatial prior map representing the log-probability of the class label at given image coordinates, and the resulting output representing the posterior probability of the class label; however, the spatial map parameters were learned (per fold) during trainging using gradient descent and may not represent true prior probabilities.

Training used the Adam optimiser minimizing a loss weighting L2 regularization and per-organ Dice scores. After training, the label with the maximum voxel-wise softmax probability was chosen. Segmentations were post-processed by eliminating, for each organ, each connected component comprising <10% of the total label volume, and upsampled to the original resolution.

Multi-atlas-based segmentation (for comparison) Our proposed algorithm is directly compared to an existing algorithm using multi-atlas registration and label fusion. First, multiple atlas images were registered to the input image using NiftyReg (`http://niftk.org/niftyreg`) to maximize normalized mutual information under affine then B-spline transformations. Then, transformed reference labels were combined using two fusion algorithms – majority voting, and joint label fusion [15] – yielding two sets of segmentations. Majority voting is a fast fusion algorithm where the segmentation labels are the
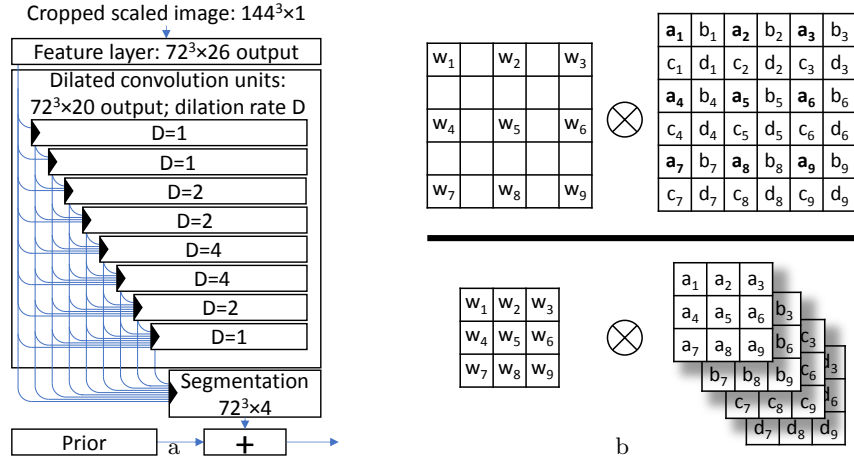
Fig. 1. a) Network architecture; b) Dilated convolutions (top) generate the same output as more efficient regular convolutions on reordered input data (bottom).

voxel-wise modes of the transformed segmentation labels, and was implemented in Matlab. Joint label fusion is a statistical fusion algorithm where the segmentation label is the weighted average of the transformed labels, with weights computed based on the local image similarity between the transformed atlas and input images, while modelling correlations between atlas images. Joint label fusion, implemented in the publicly available PICSL Multi-Atlas Segmentation Tool (`https://www.nitrc.org/projects/picsl_malf`), and its variants have yielded the highest performance in MICCAI multi-atlas labeling grand challenges in 2012 [7] and 2015. Default parameters were used except for a B-spline control point spacing of 10 mm (instead of 5 voxels) to allow for anisotropic voxels. Segmentations were post-processed as in the previous section.

Evaluation The segmentation algorithms were evaluated with an 8-fold cross-validation over all 72 subjects. In each fold, we compared the segmentation of each organ in each test image from each algorithm to the reference segmentation using 3 metrics: Dice coefficient – $2|A \cap B|/(|A| + |B|)$; symmetric mean boundary distance – $(\overline{D(A, B)} + \overline{D(B, A)})/2$; and symmetric 95% Hausdorff distance – $(percentile(D(A, B), 95) + percentile(D(B, A), 95))/2$; where A is the algorithm segmentation, B is the reference segmentation, Ω_A is the set of boundary pixels of A, and $D(A, B) = \{\min_{x \in \Omega_B} ||x - y|| \, |y \in \Omega_A\}$ is the set of boundary distances from Ω_A to Ω_B. The Dice coefficient reflects the voxel-wise overlap. The mean boundary and 95% Hausdorff distances reflect the agreement between segmentation boundaries, with the latter being more sensitive to localized disagreements.

We compared the three algorithms for each organ and metric using Friedman tests (non-parametric repeated-measures ANOVA) with Benjamini–Hochberg false-discovery rate multiple comparison correction ($\alpha = 0.05$) for pairwise tests.
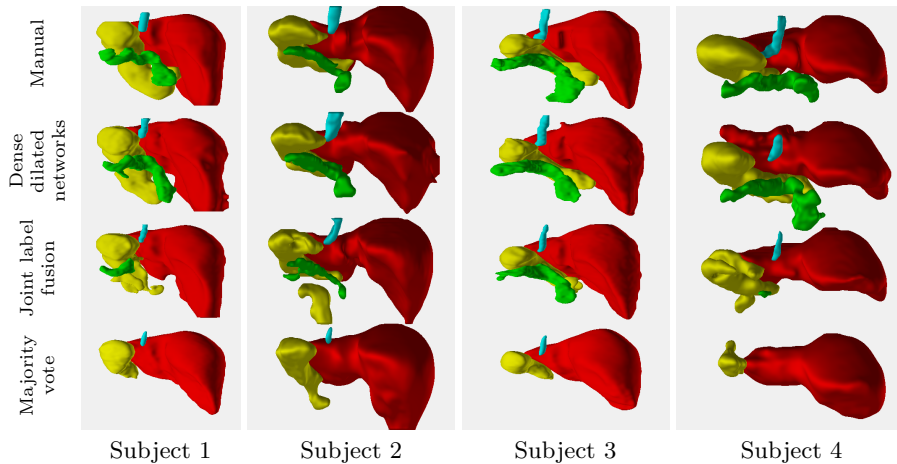
Fig. 2. Posterior view of four segmentations with Dice scores closest to the median (1 & 2), and to the 75th and 25th percentile (3 & 4). Liver (red), pancreas (green), stomach (yellow) and esophagus (cyan) segmentations were generated, from top to bottom, by manual segmentation, deep learning, joint label fusion and majority voting methods.

3 Results

The median and quartiles of the segmentation evaluation metrics are reported in Table 1, with representative segmentations shown in Fig. 2.

Majority voting consistently underperformed the other algorithms, and failed to correctly identify any voxels as pancreas in the majority of subjects. The deep-learning-based algorithm yielded more accurate segmentations than the joint-label-fusion algorithm for the smaller organs – pancreas, stomach and esophagus – showing significantly higher Dice scores for the pancreas (median 66 vs. 37), stomach (median 83 vs 72), and esophagus (median 73 vs 54) and significantly lower boundary distances for the pancreas and esophagus, as determined by the Friedman tests. Conversely, the joint-label-fusion algorithm yielded statistically significantly more accurate segmentations for the liver by all three measures, although the differences in median values were small. As seen in Fig. 2, both label fusion methods frequently under-segmented the pancreas, suggesting challenges in consistently registering this thin organ with variable abdominal position [14].

4 Discussion

This paper presents a deep-learning-based algorithm to segment liver, pancreas, stomach and esophagus on abdominal CT, while avoiding challenging inter-patient registration of abdominal organs.

Table 1. Segmentation metrics from the cross-validation (Median [first, third quartile])

	Liver	Pancreas	Stomach	Esophagus
	Dice coefficient (%)			
Majority voting	85 [78,91]	0 [0,0]	49 [23,65]	29 [4,44]
Joint label fusion	93 [90,95]	37 [10,69]	72 [52,83]	54 [39,64]
Dense dilated networks	92 [90,93]	66 [54,74]	83 [73,87]	73 [65,78]
	Mean boundary distance (mm)			
Majority voting	6.0 [3.4,8.4]	$-^1$ [27.6,$-^1$]	12.7 [6.7,22.3]	4.4 [2.7,7.4]
Joint label fusion	2.3 [1.5,3.2]	5.8 [1.6,13.8]	4.6 [2.8,9.5]	2.3 [1.6,4.2]
Dense dilated networks	2.6 [2.1,3.3]	2.8 [1.8,4.4]	3.4 [2.6,4.7]	1.4 [1.1,1.8]
	95% Hausdorff distance (mm)			
Majority voting	23.3 [13.3,31.6]	$-^1$ [56.2,$-^1$]	38.8 [21.2,54.6]	9.9 [7.3,15.6]
Joint label fusion	8.2 [5.2,14.0]	19.3 [5.4,34.0]	19.9 [10.5,32.0]	6.1 [4.6,12.6]
Dense dilated networks	10.8 [9.0,14.3]	10.8 [6.7,17.1]	13.7 [9.0,18.1]	4.7 [3.5,6.3]

[1]The boundary distance is not defined for empty segmentations.

Endoscope navigation through the gastrointestinal tract could benefit from segmentations of multiple gastrointestinal and surrounding organs. Many previous studies have proposed methods for multi-organ segmentation of abdominal CT, principally based on multi-atlas segmentation [16,9,14,13,12] or statistical shape models [2,9]. Organs surrounding the GI tract, such as the liver and pancreas, are included in many of these studies, but esophagus and stomach segmentation has received little attention, likely due to the lack of available reference segmentations. Liver segmentation has consistently yielded higher Dice scores (82–95) than other anatomy (pancreas [45–74], stomach [10–87], esophagus [36-43]). Dice scores from some previous studies are given in Table 2.

Table 2. Dice scores for previous abdominal CT multi-organ segmentatation methods. Different data sets and segmentation of unlisted organs preclude direct comparisons.

First Author	Year	Approach	Dice scores %			
			Liver	Pancreas	Stomach	Esophagus
Hu [5]	2017	deep learning	96			
Xu [16]	2015	multi-atlas	91	45	55	43
Cerrolaza [2]	2015	shape model	82	74	87	
Okada [9]	2015	shape models + multi-atlas	94	73		
Tong [14]	2015	multi-atlas	95	71		
Casiraghi [1]	2009	intensity-based	95			
Shimizu [12]	2007	multi-atlas	94	52	55	36

As in previous work, observed Dice scores were substantially higher for the liver than for the other smaller organs. While noting that quantitative metrics for algorithms evaluated on different data sets are not directly comparable, the proposed segmentation yielded stomach and esophagus segmentations with Dice

scores higher than previous studies, and liver and pancreas segmentations with Dice scores in the range observed in previous studies. Notably, compared to the two previous studies generating segmentations of all four organs segmented in this work, Dice scores for deep-learning-based esophagus, stomach and pancreas segmentations were higher, and for liver segmentations were within 2%.

The dilated convolutions with dense skip connections in our network address two key challenges in information and gradient propagation in deep convolutional networks for segmentation: (1) using both local and distant image information and (2) using both low and high-resolution image information. While large convolutional kernels enable the propagation of high-resolution local and distant image information, the larger kernels result in higher parameter counts (particularly in 3D convolution), challenging learning, and increasing the risk of over-fitting. A second approach has been to use spatial pooling or down-sampling layers such that small convolution kernels (with low parameter counts) have a large effective spatial extent followed by upsampling or transpose convolutions to regain the high resolution representation needed for segmentation. This approach limits the representation of high-resolution information through the network, motivating skip connections between early high-resolution representations and later upsampled representations. Our architecture avoids this by using high-resolution representations throughout the network and dilated convolutions to propagate high resolution information at large spatial scales.

Our conclusions should be qualified by some limitations. Algorithm parameters were not extensively optimized for this application; the reported performance of both proposed and comparison algorithms may underestimate their potential performance. The evaluation metrics measure segmentation fidelity with the manual reference, and not the clinical utility of the resulting segmentations for aiding endoscopic navigation. Future work will evaluate whether the proposed algorithm is already sufficiently accurate to provide a 3D patient-specific anatomical reference to aid endoscopic navigation. Finally, to support guidance for endoscopy in the gastrointestinal tract, segmentations of the duodenum, gallbladder, left kidney and vasculature, where sufficiently large reference segmentation data sets are not yet available, would be a valuable addition.

Dilated convolutional networks with dense skip connections can segment the liver, pancreas, stomach and esophagus in abdominal CT without image registration. Our proposed method achieved lower boundary distance errors (for pancreas and esophagus) and higher overlap (for pancreas, esophagus and stomach) with manual segmentations than a recent multi-atlas label fusion algorithm. Such automatically generated segmentations of abdominal anatomy have the potential to support image-guided navigation in pancreatobiliary endoscopy procedures.

5   Acknowledgements

References

1. Casiraghi, E., Campadelli, P., Pratissoli, S., Lombardi, G.: Automatic abdominal organ segmentation from CT images. ELCVIA 8(1) (2009)
2. Cerrolaza, J.J., Reyes, M., Summers, R.M., González-Ballester, M.Á., Linguraru, M.G.: Automatic multi-resolution shape modeling of multi-organ structures. MedIA 25(1), 11–21 (2015)
3. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (TCIA): Maintaining and operating a public information repository. Journal of Digital Imaging 26(6), 1045 (Dec 2013)
4. Eisen, G.M., Dominitz, J.A., Faigel, D.O., Goldstein, J.A., Petersen, B.T., Raddawi, H.M., Ryan, M.E., Vargo, J.J., Young, H.S., Wheeler-Harbaugh, J., et al.: Guidelines for credentialing and granting privileges for endoscopic ultrasound. Gastrointestinal endoscopy 54(6), 811–814 (2001)
5. Hu, P., Wu, F., Peng, J., Bao, Y., Chen, F., Kong, D.: Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. IJCARS pp. 1–13 (2016)
6. Huang, G., Liu, Z., Weinberger, K.Q., van der Maaten, L.: Densely connected convolutional networks. arXiv:1608.06993 (2016)
7. Landman, B., Warfield, S. (eds.): MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling (2012)
8. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: IEEE 3D Vis. pp. 565–571 (2016)
9. Okada, T., Linguraru, M.G., Hori, M., Summers, R.M., Tomiyama, N., Sato, Y.: Abdominal multi-organ segmentation from CT images using conditional shape–location and unsupervised intensity priors. MedIA 26(1), 1–18 (2015)
10. Roth, H.R., Farag, A., Turkbey, E.B., Lu, L., Liu, J., Summers, R.M.: Data from TCIA Pancreas-CT (2016)
11. Roth, H.R., Lu, L., Farag, A., Shin, H.C., Liu, J., Turkbey, E.B., Summers, R.M.: Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In: MICCAI. pp. 556–564. Springer (2015)
12. Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H., Nawano, S., Smutek, D.: Segmentation of multiple organs in non-contrast 3D abdominal CT images. IJCARS 2(3), 135–142 (2007)
13. Suzuki, M., Linguraru, M.G., Okada, K.: Multi-organ segmentation with missing organs in abdominal CT images. In: MICCAI. pp. 418–425. Springer (2012)
14. Tong, T., Wolz, R., Wang, Z., Gao, Q., Misawa, K., Fujiwara, M., Mori, K., Hajnal, J.V., Rueckert, D.: Discriminative dictionary learning for abdominal multi-organ segmentation. MedIA 23(1), 92–104 (2015)
15. Wang, H., Suh, J.W., Das, S.R., Pluta, J.B., Craige, C., Yushkevich, P.A.: Multi-atlas segmentation with joint label fusion. IEEE TPAMI 35(3), 611–623 (2013)
16. Xu, Z., Burke, R.P., Lee, C.P., Baucom, R.B., Poulose, B.K., Abramson, R.G., Landman, B.A.: Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning. MedIA 24(1), 18–27 (2015)
17. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv:1511.07122 (2015)